AD_____

Award Number: W81XWH-

TITLE:

PRINCIPAL INVESTIGATOR:

CONTRACTING ORGANIZATION: University of T

REPORT DATE: June 20FF

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release; distribution unlimited

1. REPORT DATE (DD-MM-YYYY) 01-06-2011	2. REPORT TYPE Annual	3. DATES COVERED (From - To) 1 JUN 2010 - 31 MAY 2011
4. TITLE AND SUBTITLE DEFINING GENOMIC CHANGES IN TRIPLE NEGATIVE BREAST CANCER IN WOMEN OF AFRICAN DESCENT		5a. CONTRACT NUMBER
		5b. GRANT NUMBER W81XWH-09-1-0069
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) DR. LISA BAUMBACH		5d. PROJECT NUMBER
		5e. TASK NUMBER
E-Mail: lbaumbac@med.miami.edu		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Miami Miami, FL 33136		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited		
13. SUPPLEMENTARY NOTES		

14. ABSTRACT

Abstract on next page.

15. SUBJECT TERMS

Triple negative breast cancer, Ethnic disparities, Breast cancer amongst African-Americans and Africans, Gene expression profiling, Array CGH

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U	UU	11	19b. TELEPHONE NUMBER (include area code)

14. ABSTRACT

Background: Breast cancer (BC) is the second leading cause of cancer death among African-American (AA) women, with mortality 20% greater than that in Caucasians (Cauc). However, the basis for such disparity remains an enigma. Recent observations from our laboratory suggest the involvement of unidentified genes contributing to AA BC risk. Matched tumor and normal FFPE samples from Cauc and AA patients were obtained from the UM /Sylvester Breast Tissue Bank (UM/S BTB) under an IRB-approved protocol. Based on analysis of 22,000 transcripts, ethnic specific gene expression patterns were identified that may provide important new insights into molecular mechanisms of ethnic subtype differences in clinical outcomes. We propose to extend these preliminary findings to a large African tumor bank [available via collaboration between Drs. Peter A. Bird (Kijabe, Kenya) and Mark Pegram (UM Sylvester).] Additionally, we propose to analyze chromosomal alterations associated with gene expression differences utilizing array cGH (in collaboration with Alan Ashworth, England). This work will contribute to development of rationale designs of preventive, predictive and therapeutic measures for BC in different ethnicities, and thus, a significant reduction in current ethnic-specific disparities in BC incidence, morbidity and mortality.

Hypothesis: *Discrete genomic alterations and gene expression changes will be identified and shared between triple negative tumor specimens within an ethnic group, i.e., North Americans/African decent and Kenya.* **Aim I:** *Analyze and compare genome-wide differences in gene expression in BC samples of AA ancestry vs. native African (Kijabe) samples (Drs. Pegram, Baumbach, Bird, Halsey). Aim II: Investigate possible chromosomal alterations associated with gene expression differences* (*Drs. Pegram, Baumbach, Ashworth).* **Aim III:** *Analyze ancestry of each sample using a panel of ancestry-informative DNA markers (Drs. Kittles, Baumbach).*

Synergy Statement: The proposed investigations are highly synergistic. This study will also allow for the first direct comparison of gene expression/genomic copy number data in triple negative tumor specimens across Americans of African decent and Kenyan East Africans. We will correlate all experimental data with a spectrum of clinical data available on study subjects, and apply covariate modeling and logistic regression analysis to determine possible correlations between genomic signatures, genomic changes, clinical tumor characteristics and outcomes/ response measures among and across ethnic groups.

Table of Contents

3) INTRODUCTION

The advent of microarray technology has enabled the robust, high throughput analysis of disease specific transcriptomes, including those in breast tumor specimens. Indeed, the molecular classification of breast cancer has been revolutionized by the advent of gene expression profiling. However, currently available commercial microarray design focuses on the most commonly known and characterized genes from all body tissues, therefore only a subset of genes on a generic microarray will yield informative results for any tissue-specific study. Moreover, since the transcriptome of a given tissue contains tissue/disease-specific splice variation as well as non-coding RNAs, many important transcripts solely expressed in the tissue of interest will not be represented. One innovative solution to this problem that we will utilize in this project is to exploit custom breast cancer-specific arrays developed by our collaborators at Almac Diagnostics. With tens of thousands of transcripts not found on generic arrays, specificity of differential gene expression patterns will be significantly enhanced. Furthermore, the use of expression array technology historically has been dependent upon the availability of intact RNA from fresh frozen tumor tissue for analysis, thus study of the many large retrospective cohorts with annotated clinical follow-up has not been possible. RNA extracted from FFPE samples tends to have shorter median length from 3' to 5' and the detection of these transcripts on generic array platforms is rarely successful. However, using an innovative approach we have recently successfully tested novel array probes specifically designed to detect partially degraded RNA from formalin-fixed, paraffin-embedded (FFPE) breast tumor material from samples at the University of Miami. The use of a probeset with extreme 3' sequence mitigates this previous technical limitation, and thus is considered highly innovative.

Another innovation in this study is the genomic analysis of a published East African breast cancer cohort, the largest of its kind from the region. Importantly, the integration of high density array cGH technology with the expression array data is highly innovative (to our knowledge, the first study of this kind in a native African, or even African American cohort). This approach will allow identification of ethnic specific copy number variation and loss of heterozygosity, and their relation to gene expression changes. Finally, the incorporation of an ancestry marker panel makes this a particularly novel study which is sure to produce data of interest to the community. Our eventual goal will be to develop further understanding of biology of disease, prognostic biomarkers, and eventually, the targets for therapeutics for ethnic-specific subgroups in breast cancer.

4) BODY

Characteristics of Study Population

Breast cancer is the second leading cause of cancer death among African-American (AA) women (1). Mortality is 20% greater than that in Caucasian (Cauc) women, and is partially attributed to more aggressive disease and poorer prognosis. In addition, AA women ≤ 50 years have the highest rate of new breast cancer cases in the US (1,2). General consensus exists that AA women of all ages are more likely to have poorly differentiated breast cancer, which is likely to occur at an earlier age, be ER and PR negative, and to have a higher proliferative fraction - all factors associated with more aggressive tumors (2). Therefore, the prognosis in AA patients is worse, even adjusted for stage of presentation. Ethnic-specific differences in response to adjuvant therapy have also been reported (3,4). Taken together, the cumulative data suggests that intrinsic, ethnic-specific, and biological/genetic differences contribute to disparities in breast cancer morbidity and mortality.

A recent study by Bird et al (5) focused on a cohort of BC patients from the Kijabe Hospital in Kenya and reported a very low frequency of hormone receptor expression: 24% ER-positive and 34% ER-or PR-positive tumors. Compared to breast cancer in Western or Cauc populations, the Kijabe patients have a high proportion of poorly differentiated, advanced cancers and irrespective of disease stage, were much less likely to be hormone sensitive (ER and PR negative). Overall, the possibility of inherently more aggressive tumor biology, coupled with low hormone receptor sensitivity, may represent manifestations of modified biology in African populations. This study further characterizes the tumors in a Kijabe clinical cohort. A set of 55 residual pathology tissue

blocks were obtained from Dr. Peter Bird in Kenya. These samples were selected by Dr. Bird on the basis on of being ER or PR negative by clinical testing. Her2 testing had not previously been done on any of the samples. For all 55 cases sections were cut and stained by immunohistochemistry for ER, PR and Her2. The stained slides were evaluated by UM pathology to find all cases which were triple negative (negative for ER, PR and Her2). The table below shows the results of immunohistochemical staining.

Table 1. Hormone Receptor Status of Kijabe Breast Cancer Cohort

Receptor Status	No. of Samples
ER-/ PR-/ Her2-	31
ER-/ PR-/ Her2+	10
ER-/ PR+/ Her2-	0
ER+/ PR-/ Her2-	8
ER+/ PR-/ Her2+	2
ER+/ PR+/ Her2-	2
ER+/ PR+/ Her2+	2

Staining showed that 31 out of the 55 samples (or 56%) were triple negative breast cancer samples. These 31 samples were selected for use in this project. Of the remaining cases 12 were positive for Her2 staining with nine samples (or 16%) with strongly positive score of +3. Her2 staining is interpreted on the maximum area of staining intensity as follows: 0 = no staining; +1 = weak, incomplete membranous staining; +2 = moderate, complete membranous staining of at least 10% of invasive tumor cells; and +3 = strong membranous staining of at least 10% of invasive tumor cells. Cases interpreted as 0 or +1 are considered negative, and cases interpreted as +2 or +3 are considered positive. Many of the cases with Her2 +3 staining appeared to be advance stage aggressive cancers. In general the cohort of 55 samples reflects the overall advanced stage and high incidence of hormone receptor negative cases seen in both African and African-American breast cancer cases.

Gene Expression Array Studies
The 31 triple negative Kijabe samples were then matched with an equal number of African-American samples from South Florida. Samples were obtained from residual pathology tissue blocks for cases confirmed to be triple negative by immunohistochemistry. Once samples are identified RNA is extracted and expression profiling done using the Almac Diagnositics Breast Cancer Disease Specific Array (DSA). Quality control checks are completed at each step of the process, for RNA quality spectrophotometer and the Agilent Bioanalyzer are used, the Bioanalyzer provides more sensitive qualitative analysis from less RNA than other traditional methods. The bioanalyzer uses a fluorescent assay and electrophoretic separation to evaluate RNA samples qualitatively. The software creates a graph called an electropherogram, high quality RNA electropherograms exhibit two primary characteristics. First, clear 28S and 18S peaks and secondly, there should be low noise between the peaks and minimal low molecular weight contamination. Samples meeting these criteria are then processed for hybridization to the DSA array.

Data resulting from the DSA hybridizations are checked for quality control by first looking at the distribution of thesample data (histogram of normalised intensity values) will be assessed to determine what statistical tests will be applied in later stages of analysis. The data had a normal distribution so K-Means Clustering was performed. In K-Means clustering groups are created which shows the relationships among the expression levels of conditions or samples. This allows identification any spurious samples, a particular concern when replicates are included in an experiment. K-Means is used because of prior knowledge of samples condition as being either from tumor or normal tissue.

In addition to K-Means Clustering, Principal Components Analysis is also performed: This is a decomposition technique that produces a set of expression patterns known as principal components. Linear combinations of these patterns can be assembled to represent the behavior of all of the genes in a given data set. Although not a clustering technique, the aim of PCA is similar to that of clustering. It is a tool to characterize the most abundant themes or building blocks that reoccur in many genes in the experiment.

Two-dimensional hierarchical clustering analysis was performed to examine the gene expression patterns across samples groups at intensity level. The result was shown in a heatmap (see example in Figure 1 below). From the heatmap, data from the Kenyan (Native African) tumor samples are clearly different in expression pattern from adjacent normal tissue from African-Americans.

Figure 1. Gene Expression Pattern of Native African Triple Negative Breast Cancer
and African-American Adjacent Normal Breast Tissue
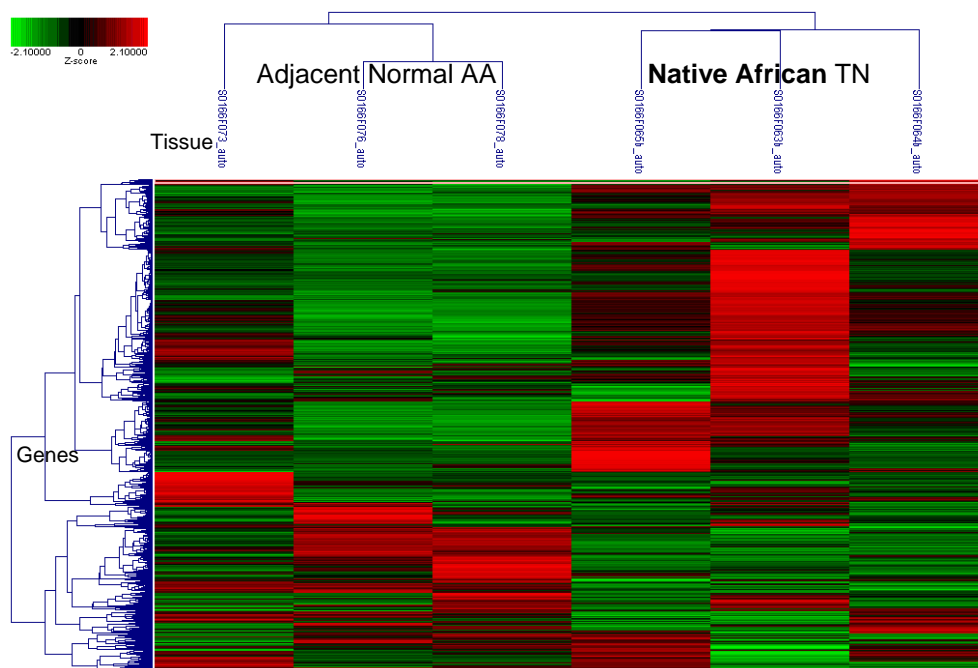
Also, following quality control measures, Stringent and Less Stringent Gene Lists are generated from the expression data. For the differentially expressed genes, genes with intensity greater than the background intensity plus the 3 standard deviations are retained as presence call in the data. For stringent genes, genes with intensity greater than 2X background intensity were retained in the sample group. For the Differentially Expressed Gene List, cut-offs of a p-value of 0.01 in 2way ANOVA and paired-t-tests are applied. Sequences with significant statistical confidence (p-value<0.01 in both tests) were retained in these differentially expressed gene lists. These genes/transcripts are subjected to pathway analysis in Metacore GeneGo program.

Copy Number Variation (aCGH) arrays

As cancer cells develop, they undergo dramatic DNA rearrangements such as chromosome loss (Loss of Heterozygosity; LOH) or duplication or translocation. We are using high density CGH arrays to analyze genome wide variation to assess whether gene expression differences may be due to chromosomal alterations.

aCGH is performed using the Breakthrough Breast Cancer 32K tiling path microarray platform, which has a complete coverage of the whole genome with a resolution of 50kb. Details of labelling, hybridization, washes, image acquisition, data pre-processing, normalization and analysis were previously reported (6). Data analysis of these arrays is as follows, cases with >10% of clones missing and clones for which data are not available in ≥10% of cases will be excluded. Log2 ratios will be normalized for spatial and intensity dependent biases using a two-

dimensional loess regression followed by a BAC-dependent bias correction (6). The final dataset of BAC clones with unambiguous mapping information according to the build hg19 of the human genome is used for further analysis. A categorical analysis is applied to the BACs after classifying them as representing gain, loss, or no-change according to their smoothed Log2 ratio values. Threshold values have already been defined in previous studies (6). These thresholds accurately identify low level gains, which are defined as a smoothed Log2 ratio of between 0.12 and 0.45, corresponding to approximately 3-4 copies of the locus, whilst gene amplifications are defined as having a Log2 ratio > 0.45, corresponding to more than 5 copies (see Figure 2 below for example aCGH results).
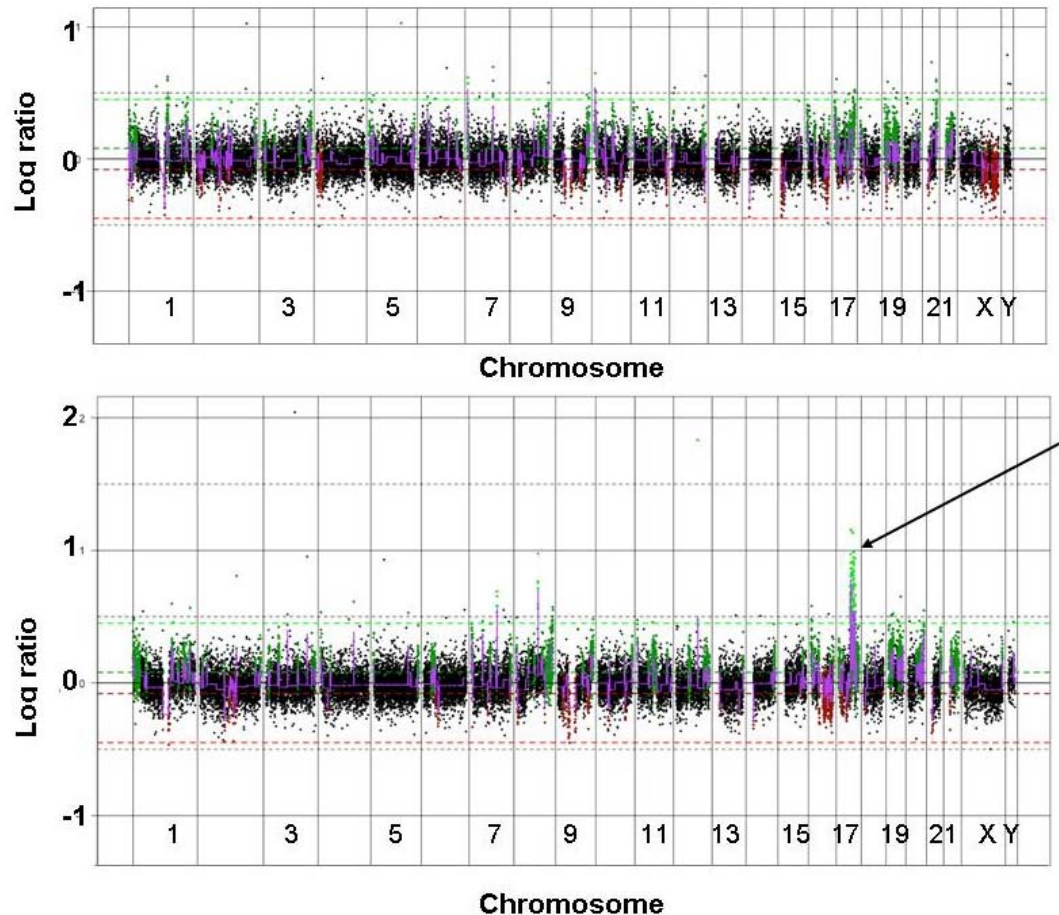

Figure 2. Copy number changes in two Kijabee Kenyan samples from project cohort. Genome plots from aCGH results with log2 ratios for each clone (*Y axis*) plotted according to chromosomal location (*X axis*). *Horizontal line,* centromere. *Green,* gains; *red,* losses. *Top Panel:* sample 1886, showing no large-scale changes in DNA copy number. *Bottom Panel:* sample 1887, showing a single large alteration, a duplication in chromosome 17 (see arrow).

**Ancestry Informative Markers (AIMS)**

This set of genome-spanning SNPs provides a rich source of information for examining admixture in African Americans these are used rule out spurious results due to underlying population stratification. These portion of the project is to genotype 100 carefully selected ancestry informative markers for all the AA samples. 100 autosomal SNP AIMs are genotyped using the Sequenom MassARRAY platform and iPLEX™chemistry. iPLEX assays were designed utilizing the Sequenom Assay Design software, allowing for single base extension (SBE) designs used for multiplexing. Individual SNP genotype calls are then generated using Sequenom TYPER software, which automatically calls allele specific peaks according to their expected masses. Quality control checks include genotyping in duplicate multiple samples (10%) in each plate of DNA; cases and unaffected controls are gridded together in each plate to avoid any systematic biases between plates. Individual African ancestry will be estimated

from the genotype data using the Bayesian Markov Chain-Monte Carlo (MCMC) method implemented in the program STRUCTURE version 2.1. STRUCTURE will be run under the admixture model using prior population information and independent allele frequencies. Ancestry estimates generated from these AIMs will allow for accurate estimates of European ancestries in our AA subjects, allowing us to utilize individual ancestry estimates as additional covariates in overall experimental analyses. Data for all of the African-American samples in this project has not been completed but data from a separate control sample set of 112 African-American samples for South Florida has been completed and shows a range of 62%-98% African ancestry with a mean of approximately 72% African ancestry. This cohort was completed in conjunction with another breast cancer project represents a random sample of South Florida African-Americans and should reflect the overall range of African ancestry in the African-American samples used in the gene expression/CNV studies in this current project.

Analysis and comparison of data for genome expression profiling/copy number variation (CNV) data and ancestry determination.
Once each portion of the project is completed we will begin to compare data obtained from gene expression array with that from copy number variation to determine if the differences in gene expression might be explained by the loss or gain of chromosomal segments within the tumor DNA. We will also correlate the gene expression differences with the results of the ancestry informative markers to assess the influence of ancestry upon the gene expression results of the African-American samples.

Timeline for completion of work: Illustrated below is the projected timeline for completion of all tasks in the study given the one year no-cost extension granted last month.
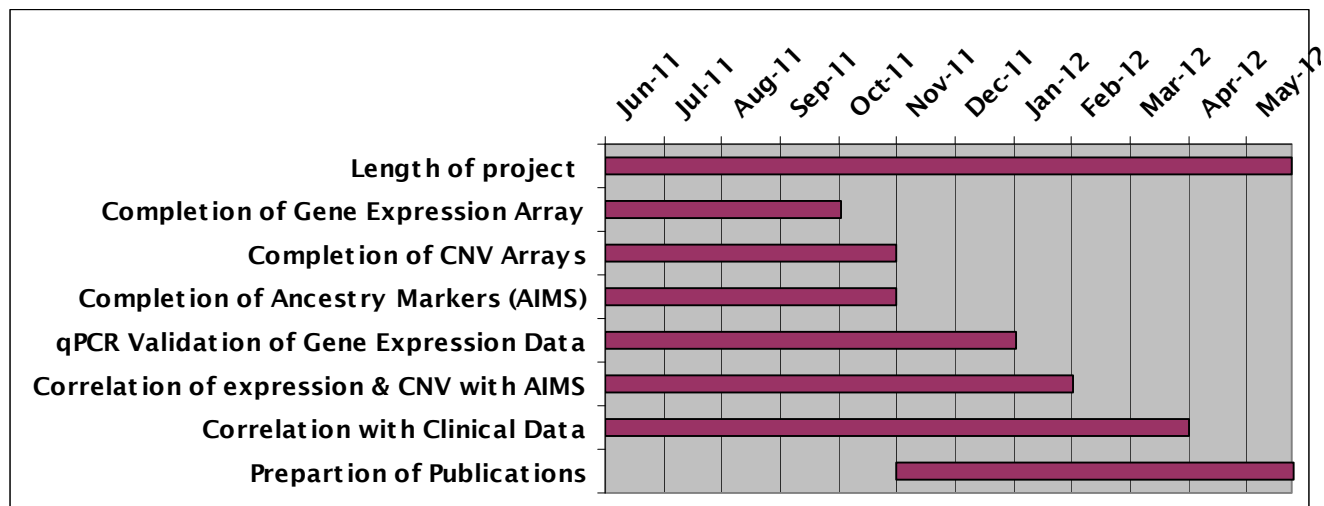
5) KEY RESEARCH ACCOMPLISHMENTS

- Task 1: Determination of HER2 status in the Kijabe clinical cohort using immunohistochemitry.

- Task 2: Extraction and preparation of DNA and RNA from FFPE tumor samples from North American African, African African, and Caucasian cohorts.

- Task 3: Aim I, Analyze and compare genome-wide transcript expression in BC samples of AA ancestry vs. native African (Kijabe) samples.

- Task 4: Aim II, Investigate possible chromosomal alterations associated with gene expression differences using array cGH.

6) REPORTABLE OUTCOMES

- **Continued Identification of Ethnic Specific Differences in Breast Tissue – on the Road to Biomarker Discovery in Breast Cancer.** Lisa L. Baumbach-Reardon, Mary Ellen Ahearn, Carmen Gomez, Aldo Mejias, Merce Jorda, Tom Halsey, Jim Yan, Kevin Ellison, Karl Mulligan, Mark Pegram. Univ. of Miami Medical School, Miami, FL, Almac Diagnostics, Durham, NC. *AACR Special Conference*, *The Future of Molecular Epidemiology: New Tools, Biomarkers, and Opportunities*, June 6 – June 9, 2010, Miami, Florida.

- **Genomic/genetic differences in breast cancer across ethnicities.** Lisa L. Baumbach-Reardon. University of Miami **Platform Presentation/ Invited Speaker; AACR** *Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved,* September 30-October 3, 2010.

- **Continued Identification of Ethnic Specific Differences in Breast Cancer and Normal Breast Tissue.** L. Baumbach, M. E. Ahearn, C. Gomez, A. Mejias, M. Jorda, T. Halsey, J. Yan, K. Ellison, K. Mulligan, R. Kittles, A. Ashworth, M. Pegram Univ Miami School of Medicine, Miami, FL; Almac Diagnostics, Durham, NC; University of Illinois at Chicago, Chicago, IL; Breakthrough Cancer Research Center, London, UK. American *Society of Human Genetics Annual Meeting* November 2010.

- **Gene Expression Profiling of Formalin-Fixed, Paraffin-Embedded (FFPE) Breast Cancer Samples and Analysis of Intrinsic Subtypes** Baumbach LL, Gomez C, Yan J, Halsey T, Ahearn ME, Jorda M, Kennedy R, ODonnel J, McDyer F, Deharo S, Pegram M. University of Miami Medical School, FL; Almac Diagnostics, Durham, NC**.** *San Antonio Breast Cancer Symposium* December 2010).

- **Identification of ethnic specific differences in breast cancer and normal breast tissue** Lisa L. Baumbach, Carmen Gomez, Jim Yan, Tom Halsey, Mary Ellen Ahearn, Merce Jorda, Mark Pegram. University of Miami School of Medicine, Miami, FL; Almac Diagnostics, Durham, NC American *Association for Cancer Research Annual Meeting (highlighted for media attention)*, Orlando, FL (2011).

- **Gene Expression Profiling of Formalin-fixed, Pariffin-embedded (FFPE) Tissues from Triple-negative Breast Cancer Patients** Baumbach, et al., (2011, manuscript in preparation).

7) CONCLUSION

RNA and DNA extracted from these samples are usually degraded, contaminated and of low quality in general. Despite the large banks of FFPE samples available for retrospective studies that include follow-up analysis of patient outcome, most of these studies currently focus on frozen samples because of the limited options available for paraffin samples. Additionally, FFPE processing holds advantages for tissue storage during prospective studies, in which many biopsies are collected but only a fraction of them are applied to downstream assays with selection based on clinical outcome. Because of the difficulty and time required to obtain fresh frozen tumor samples from the triple negative breast cancer patients with matched clinical criteria and curation, this study explored the possibility to profile both gene expression and genotype from FFPE tumor tissues. This study attempts to test and establish the feasibility and outline guidelines for selection of technology platforms and QC criteria for FFPE samples. FFPE RNA and DNA that are applied to the Almac Diagnostic Breast Cancer DSA arrays may still vary in quality and therefore require careful and rigorous QC to select samples that meet the quality standard including chip CQ and sample integrity check at profiling level. In CNV data, although the QC performance of FFPE sample are not comparable to fresh frozen sample, with careful QC and data analysis, valuable information such as LOH, and copy number assessment can still be obtained. The power comes when combine both the gene expression data with the genetic variation results; we could identify tumor suppressor genes that showed in both chromosomal aberration and transcriptional changes. These results outline guidelines for the application of FFPE samples to the same genome-wide platform already available to high-quality DNA samples, thus enabling widespread retrospective and prospective analysis of tumor samples in their most common form of storage.

8) REFERENCE

1. American Cancer Society (2007) Cancer Facts and Figures for African-Americans.
2. Chlebowski RT et al (2005) Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *JNCI* 97:6, 439-48.
3. Albain KS et al (2003) Outcomes of African Americans in SWOG BC therapy trials.*Br Can Res Treat* 77:A21.
4 Shen Y et al (2007) Are there racial differences in breast cancer treatments and clinical outcomes for women treated at M. D. Anderson Cancer Center? *Breast Cancer Res Treat* 102:347–356.
5. Bird PA, Hill AG, Houssami N. (2008) Poor hormone receptor expression in East African Breast Cancer: Evidence of a Biologically Different Disease? *Annals of Surgical Oncology*
6. Arriola E, Marchio C, Tan DS, et al. (2008) Genomic analysis of the HER2/TOP2A amplicon in breast cancer and breast cancer cell lines*. Lab Invest* 88: 491-503.

9) APPENDICES -- none